Logistic Regression Algorithm(single variable)

Juan David Rodríguez Cuervo

September 26, 2017

The **Logistic Regression Algorithm** is a Machine learning technique used for classification. It is a binary decision curve, which means that for some input x (in the *one-variable* case), it returns a value between 0-1 representing the probability of belonging to the set.

The base function for the logistic regression is the sigmoid function, which has the form

$$f(x) = \frac{1}{1 + e^{-(ax+b)}} \tag{1}$$

Different from the Linear Regression function, this is not linear. That means that the error function can't be linear either. As the distribution of the dataset points is binary, it follows the general form for binomial distribution. Thus, we get:

$$Error = E(a,b) = \sum_{i=0}^{N} f^y (1-f)^{1-y} \tag{2}$$

or what is equivalent:

$$E(a,b) = \sum_{i=0}^{N} \frac{e^{(y-1)(ax+b)}}{1 + e^{-(ax+b)}} \tag{3}$$

Both forms of the error value represent two expressions of the same object. Equation (2), however, is computationally more desirable, as it is more compact and clear. On the other hand, (3) is used better when calculating the gradients for the coefficients, as it is more explicit.

Having the error function, we can have an expression for the gradient to have a value of change for each learning step. We proceed to find those gradients on which the value variation is depending on:

$$\frac{\partial E(a,b)}{\partial a} = \sum_{i=0}^{N} x_i f(x_i)[y_i - f(x_i)] \left[e^{-(ax_i+b)}\right]^{1-y_i} \tag{4}$$

$$\frac{\partial E(a,b)}{\partial b} = \sum_{i=0}^{N} f(x_i)[y_i - f(x_i)] \left[e^{-(ax_i+b)}\right]^{1-y_i} \tag{5}$$

With (4) and (5) we solve the problem of the direction and magnitude of the movement. To avoid the divergence of the *learning process*, in the code we define a *learning_rate* which will control the movement of the values. now, for the setting of the new values, let p be the step number in which the system is at certain point. That way, we have, for each step:

$$a_{p+1} = a_p - \frac{\partial E}{\partial a_p} \times learning\_rate \tag{6}$$

$$b_{p+1} = b_p - \frac{\partial E}{\partial b_p} \times learning\_rate \tag{7}$$